EFFICIENT MAXIMUM LIKELIHOOD ESTIMATION OF A 2-D COMPLEX SINUSOIDAL BASED ON BARYCENTRIC INTERPOLATION

J. Selva

Dpt. of Physics, Systems Engineering, and Signal Theory (DFISTS)
University of Alicante, P.O. Box 99,
03080 Alicante, Spain. E-mail: Jesus.Selva@ua.es

ABSTRACT

This paper presents an efficient method to compute the maximum likelihood (ML) estimation of the parameters of a complex 2-D sinusoidal, with the complexity order of the FFT. The method is based on an accurate barycentric formula for interpolating band-limited signals, and on the fact that the ML cost function can be viewed as a signal of this type, if the time and frequency variables are switched. The method consists in first computing the DFT of the data samples, and then locating the maximum of the cost function by means of Newton's algorithm. The fact is that the complexity of the latter step is small and independent of the data size, since it makes use of the barycentric formula for obtaining the values of the cost function and its derivatives. Thus, the total complexity order is that of the FFT. The method is validated in a numerical example.

Index Terms— Frequency estimation, parameter estimation, fast Fourier transform (FFT), barycentric interpolation, sampling.

1. INTRODUCTION

The estimation of the parameters of a complex sinusoidal in either one or two dimensions is a pervasive problem in signal processing. For this problem, the maximum likelihood (ML) principle provides an optimal estimator, in the sense that it achieves the Cramer-Rao bound at relatively low signal-to-noise ratios. Yet in practice this estimator is deemed too complex, due to the associated maximization problem. In rough terms, the situation is that, though the ML cost function can be regularly sampled using the FFT efficiently, the localization of its maximum requires some search procedure, and here is where the high computational burden seems unavoidable. This observation was first stated by B. G. Quinn in [1], where it was shown that a Gauss-Newton iteration may fail to find the global maximum of the cost function, if the initial iteration is taken from the DFT of the data samples (without zero padding). This has led several researches to give up the ML approach, and look for sub-optimal estimators with lower computational burden, [2, 3]. However, in some references [4, 5, 6] there has been an attempt to recover the initial ML approach, based on two arguments. The first is that it is feasible to approximately locate the maximum of the ML cost function,

simply by looking for the DFT sample with largest module. Besides, the accuracy of this coarse localization can be improved by zero padding. And the second is that it is possible to interpolate the cost function close to this location from the nearby DFT samples with some accuracy, since it is a smooth function. These arguments, if properly exploited, make it possible to improve the performance significantly, with a complexity similar to that of the FFT. The purpose of this paper is to go one step further in this direction, and show that it is feasible to perform this interpolation with high accuracy from a small number of DFT samples, so that the actual ML estimate can be obtained with the complexity of a single FFT. The key lies in viewing the DFT as a band-limited signal in the frequency variable.

The basic concept in this paper is first presented in the next section in the context of time-domain interpolation. It is a simple an efficient technique to interpolate a band-limited signal from a small number of samples with high accuracy. Then, the estimation problem in one dimension is introduced in Sec. 3, where it is shown that the method in the next section is actually the key for an efficient solution, if the independent variable is properly interpreted. Afterward, Sec. 4 presents the extension of the method in Sec. 3 to the two-dimensional case, and finally Sec. 5 contains a numerical example.

2. HIGH ACCURACY INTERPOLATION OF A BAND-LIMITED SIGNAL AND ITS DERIVATIVES

Given a band-limited signal $s(t)$ with two-sided bandwidth B, a usual task is to interpolate its value from a finite set of samples surrounding t. If the sampling period is T with $BT < 1$ (Nyquist condition), and $2P + 1$ samples are taken symmetrically around t, then this task consists in finding a set of coefficients $a_p(u)$ such that the formula

$$s(t) \approx \sum_{p=-P}^{P} s((n-p)T)a_p(u) \tag{1}$$

is accurate, where $t = nT + u$ is the modulo-T decomposition of t,

$$n \equiv \lfloor t/T + 1/2 \rfloor, \quad u \equiv t - nT. \tag{2}$$

For fixed u, the $a_p(u)$ can be obtained numerically using filter optimization techniques [7]. Yet this approach is cumbersome, if not only $s(t)$ but also its derivatives must be interpolated for varying values of u efficiently. Recently, a so-called barycentric interpolator was derived in [8] that solves this problem satisfactorily. This interpolator takes the form

$$s(nT + u) \approx \sum_{p=-P}^{P} \frac{s((n-p)T)w_p}{u - pT} \left/ \sum_{p=-P}^{P} \frac{w_p}{u - pT} \right., \tag{3}$$

where w_p is a set of constants which are samples of a fixed function $w(t)$, $[w_p \equiv w(pT)]$. In [8], this function is given by the formula

$$w(t) \equiv \frac{\Gamma(t/T + P + 1)\Gamma(-t/T + P + 1)g(t)}{L'(t)}, \qquad (4)$$

where $\Gamma(\cdot)$ is the Gamma function, $g(t)$ is the pulse

$$\frac{\operatorname{sinc}((1 - BT)\sqrt{(t/T)^2 - (P+1)^2})}{\operatorname{sinc}(j(1 - BT)(P + 1))}, \qquad (5)$$

and $L(t)$ is the polynomial

$$L(t) \equiv \prod_{p=-P}^{P} (t - pT). \qquad (6)$$

(See [8] for further details.) The fact is that the error of (1) decreases exponentially with trend $e^{-\pi(1-BT)P}$. In practice this means that a small P is enough to obtain high accuracy. Besides, as shown in [8], Eq. (1) can be differentiated with low complexity, so as to interpolate the differentials of $s(t)$ of any order, and (3) can be evaluated using only one division. This interpolator will be the fundamental tool in the next section.

3. ML FREQUENCY ESTIMATION: 1-D CASE

Consider a signal $z(t)$ consisting of an undamped exponential with complex amplitude a and frequency f_o, contaminated by complex white noise $n(t)$ of zero mean and variance σ^2,

$$z(t) = e^{j2\pi f_o t} + n(t). \qquad (7)$$

For simplicity, f_o is assumed to lie in $[-1/2, 1/2[$. Next, assume that M samples of $z(t)$ are taken at instants $m_1, m_1 + 1, \ldots, m_1 + M - 1$ with $m_1 \equiv -\lceil M/2 \rceil$. The maximum likelihood estimation of f_o from these samples is the argument that maximizes the cost function

$$L(f) \equiv |c(f)|^2, \qquad (8)$$

where $c(f)$ is the correlation

$$c(f) \equiv \sum_{m=m_1}^{m_1+M-1} z(m)e^{-j2\pi fm}. \qquad (9)$$

Since $c(f)$ is the DFT of the samples $z(m)$, the maximum of $L(f)$ can be approximately located by selecting a frequency spacing Δf such that $1/\Delta f$ is a power of two, and then computing the samples

$$c(k\Delta f) \equiv \sum_{m=m_1}^{m_1+M-1} z(m)e^{-j2\pi km\Delta f}, \qquad (10)$$

by means of a radix-2 FFT algorithm. The cost of this operation is only $O(M \log M)$. This way, it is possible to obtain a frequency $k_o\Delta f$ that lies close to the true abscissa of the maximum of $L(f)$. However, if $\hat{f}_o$ is this abscissa, the approximation of $\hat{f}_o$ using $k_o\Delta f$ is very inaccurate. In this situation, the accuracy can be improved by reducing Δf (over-sampling), but then the computational burden becomes high.

The barycentric interpolator in the previous section provides an efficient method to obtain $\hat{f}_o$. The key idea is that the correlation $c(f)$ in (9) is a band-limited signal *in the f variable* of bandwidth $2m_1$, i.e, the variable f in (9) plays the same role as the variable

t in (1). Also, $L(f)$ is band-limited with bandwidth $4m_1$. So, the method of sampling the correlation using the FFT, and then looking for its maximum module can be reformulated taking into account these information. First, since $L(f)$ has two-sided bandwidth $4m_1$, it is necessary to sample this function with a spacing smaller than its Nyquist period $1/(4m_1)$, in order to coarsely locate its maximum. This implies that a factor-two zero padding is enough to ensure the localization of the maximum of $L(f)$. And second, since $c(f)$ has bandwidth $2m_1$, it can be interpolated using the barycentric formula in (1) with $2m_1$ in place of B, and Δf in place of T, i.e,

$$c(f) \approx \sum_{p=-P}^{P} \frac{c((k-p)\Delta f)w_p}{\gamma - p\Delta f} \bigg/ \sum_{p=-P}^{P} \frac{w_p}{\gamma - p\Delta f}, \qquad (11)$$

where $f = k\Delta f + \gamma$ is the modulo-Δf decomposition of f defined as in (2). If $\tilde{c}(f)$ denotes the approximation to $c(f)$ in (11), then one may replace the problem of maximizing $\tilde{L}(f)$ with that of maximizing

$$\tilde{L}(f) \equiv |\tilde{c}(f)|^2 \qquad (12)$$

Besides, the maximum of this function is close to the abscissa of the largest FFT sample in (10), $k_o\Delta f$.

Since P is small and there is a coarse estimate of the maximum abscissa, the maximization of (12) is a low-complexity problem that can be solved using standard numerical methods. Given that the differentials of the barycentric interpolator can be easily computed [8, Sec. IV], a suitable numerical method is Newton's algorithm, in which the rth iteration f_r is refined using

$$f_{r+1} = f_r + \tilde{L}'(f_r)/\tilde{L}''(f_r) \qquad (13)$$

where

$$\begin{aligned}
\tilde{L}'(f) &= \frac{2}{M}\operatorname{Re}\{\tilde{c}'(f)\tilde{c}(f)^*\}, \\
\tilde{L}''(f) &= \frac{2}{M}(\operatorname{Re}\{\tilde{c}''(f)\tilde{c}(f)^*\} + |\tilde{c}'(f)|^2).
\end{aligned} \qquad (14)$$

This process can be initiated with $f_1 = k_o\Delta f$, and a small number of iterations (3 to 5) is enough to obtain $\hat{f}_o$.

The computational burden of this method is given by that of the FFT, i.e, it is $O(M \log M)$, and it yields the actual ML estimate of f_o.

4. ML FREQUENCY ESTIMATION: 2-D CASE

The method in the previous section can be extended to two-dimensional signals with non-essential changes. For this, it is only necessary to consider two variables, t_1 and t_2, and repeat the same interpolation procedure. The initial model, equivalent to (7), is

$$z(t_1, t_2) = e^{j2\pi(f_{1,o}t_1 + f_{2,o}t_2)} + n(t_1, t_2). \qquad (15)$$

Next, this signal is sampled at the integer pairs (m, n) for $m_1 \leq m < m_1 + M$ and $n_1 \leq n < n_1 + N$, with $m_1 \equiv -\lceil M/2 \rceil$, $n_1 \equiv -\lceil N/2 \rceil$. The 2-D equivalent of the cost function in (8) is

$$L(f_1, f_2) \equiv |c(f_1, f_2)|^2, \qquad (16)$$

where $c(f_1, f_2)$ is the correlation

$$c(f_1, f_2) \equiv \sum_{m=m_1}^{m_1+M-1} \sum_{n=n_1}^{n_1+N-1} z(m,n)e^{-j2\pi(f_1 m + f_2 n)}. \qquad (17)$$

This function can be sampled with spacings Δf_1, Δf_2 using a radix-2 2-D FFT. These samples provide a frequency pair $(k_{1,o}\Delta f_1, k_{2,o}\Delta f_2)$ that lies close to the maximum of $\mathrm{L}(f_1, f_2)$. Then, it is possible to set up an interpolation formula like (11) but in two dimensions,

$$
\begin{aligned}
\mathrm{c}(f_1, f_2) \approx & \sum_{p_1=-P_1}^{P_1} \sum_{p_2=-P_2}^{P_2} \\
& \frac{\mathrm{c}((k_1-p_1)\Delta f_1, (k_2-p_2)\Delta f_2)w_{1,p_1}w_{2,p_2}}{(\gamma_1-p_1\Delta f_1)(\gamma_1-p_1\Delta f_1)} . \\
& \Big[\sum_{p_1=-P_1}^{P_1} \frac{w_{1,p_1}}{\gamma_1-p_1\Delta f_1} \sum_{p_2=-P_2}^{P_2} \frac{w_{2,p_2}}{\gamma_2-p_2\Delta f_2} \Big]^{-1},
\end{aligned}
\tag{18}
$$

where $f_1 = k_1\Delta f_1 + \gamma_1$ and $f_2 = k_2\Delta f_2 + \gamma_2$ are the modulo decompositions, and w_{1,p_1} and w_{2,p_2} are the barycentric weights corresponding to bandwidths $2m_1$ and $2n_1$, and truncation indices P_1 and P_2, respectively. If $\tilde{\mathrm{c}}(f_1, f_2)$ denotes the formula in (18), the problem of maximizing $\mathrm{L}(f_1, f_2)$ can be substituted by the problem of maximizing

$$
\tilde{\mathrm{L}}(f_1, f_2) \equiv |\tilde{\mathrm{c}}(f_1, f_2)|^2 .
\tag{19}
$$

Finally, the Newton iteration for this problem is

$$
(f_{1,r+1}, f_{2,r+1}) = (f_{1,r}, f_{2,r}) + \mathcal{H}\tilde{\mathrm{L}}^{-1}\nabla\tilde{\mathrm{L}},
\tag{20}
$$

where $\nabla\tilde{\mathrm{L}}$ and $\mathcal{H}\tilde{\mathrm{L}}$ are the gradient and Hessian of $\tilde{\mathrm{L}}$ respectively, evaluated at $(f_{1,r}, f_{2,r})$. These functionals are

$$
\nabla\tilde{\mathrm{L}} = \frac{2}{MN}\,\mathrm{Re}\{\nabla\tilde{\mathrm{c}}\tilde{\mathrm{c}}^*\}
\tag{21}
$$

and

$$
\mathcal{H}\tilde{\mathrm{L}} = \frac{2}{MN}\,\mathrm{Re}\{\mathcal{H}\tilde{\mathrm{c}}\,\tilde{\mathrm{c}}^* + \nabla\tilde{\mathrm{c}}\nabla\tilde{\mathrm{c}}^{\mathrm{H}}\},
\tag{22}
$$

where $\nabla\tilde{\mathrm{c}}$ and $\mathcal{H}\tilde{\mathrm{c}}$ are the corresponding gradient and Hessian of $\tilde{\mathrm{c}}$.

Since the evaluation of $\tilde{\mathrm{c}}(f_1, f_2)$ has a small cost which is independent of either M or N, the complexity is given by the 2-D FFT, i.e, it is $\mathrm{O}(MN\log(MN))$. Note that sub-optimal methods like that in [3] have complexity $\mathrm{O}(MN(M + N))$, that is, the method in this section yields the ML estimate and, besides, has a smaller complexity order.

5. NUMERICAL EXAMPLES

Let $\phi(f, t)$ denote the value delivered by the barycentric formula in (3), when the input signal is the undamped exponential $\mathrm{s}(t) = e^{j2\pi ft}$. Since the functions to interpolate in this paper are sums of exponentials like $e^{j2\pi ft}$, a simple way to assess the interpolation accuracy is to evaluate the maximum module of $e^{j2\pi fu} - \phi(f, u)$, for $|u| \leq T/2$ and f varying in $[-B/2, B/2]$, i.e, to assess the spectrum function

$$
\mathrm{E}(f) \equiv \max_{|u|\leq T/2} |e^{j2\pi fu} - \phi(f, u)| .
\tag{23}
$$

Fig 1 shows $\mathrm{E}(f)$ for $BT = 0.25$ and several truncation indices P. Note that any accuracy can be achieved uniformly in $[-B/2, B/2]$, by slightly increasing P.

In order to test the interpolation error in a specific example, a 2-D undamped exponential with $M = 500$, $N = 651$ and frequencies $f_{o,1} = 0.234452$ and $f_{o,2} = -0.143254$ was generated. Then, complex white noise was added, so that the signal-to-noise ratio was
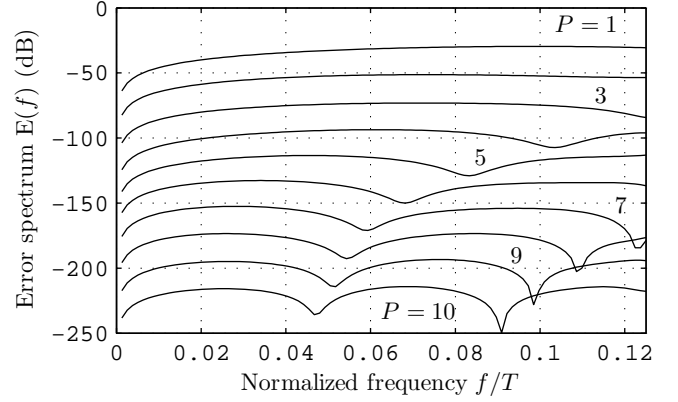

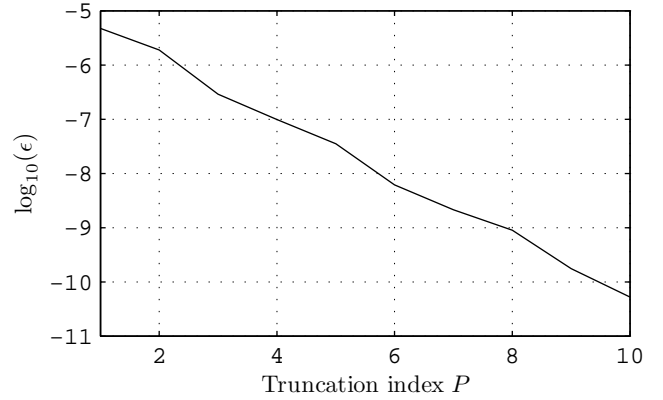**Fig. 1**. Error spectrum $\mathrm{E}(f)$ of the barycentric interpolator in (3) for several truncation indices P.

Fig. 2. Truncation index P versus the $\log_{10}$ of the maximum interpolation error ($\log_{10}\epsilon$) for the cost function $\mathrm{L}(f_1, f_2)$.

$S/N = 5$ dB. Fig. 2 shows the interpolation error for the cost function $\mathrm{L}(f_1, f_2)$, defined by

$$
\epsilon \equiv \frac{\max_{f_1,f_2} |\mathrm{L}(f_1, f_2) - \tilde{\mathrm{L}}(f_1, f_2)|}{\max_{f_1,f_2} \mathrm{L}(f_1, f_2)},
\tag{24}
$$

for varying truncation index P. Again, it is clear that any accuracy can be achieved by slightly increasing P.

Next, two estimators were compared in this example. To describe them, let $\mathbf{Z}$ denote the data matrix obtained by sampling (15) as described in that section,

$$
[\mathbf{Z}]_{m,n} \equiv \mathrm{z}((m_1 + m - 1)T_1, (n_1 + n - 1)T_2),
\tag{25}
$$

and let $\mathbf{u}_1$, $\mathbf{v}_1$ denote its first left and right singular vectors, respectively. The first method consisted in computing the 1-D ML estimator from $\mathbf{u}_1$ and $\mathbf{v}_1$ so as to obtain the respective estimations of $f_{o,1}$ and $f_{o,2}$,

$$
\begin{aligned}
\hat{f}_{o,1} &= \arg\max_f \Big| \sum_{m=1}^{M} [\mathbf{u}_1]_m e^{-j2\pi(m_1-1+m)f} \Big|^2 . \\
\hat{f}_{o,2} &= \arg\max_f \Big| \sum_{n=1}^{N} [\mathbf{v}_1]_n e^{-j2\pi(n_1-1+n)f} \Big|^2 .
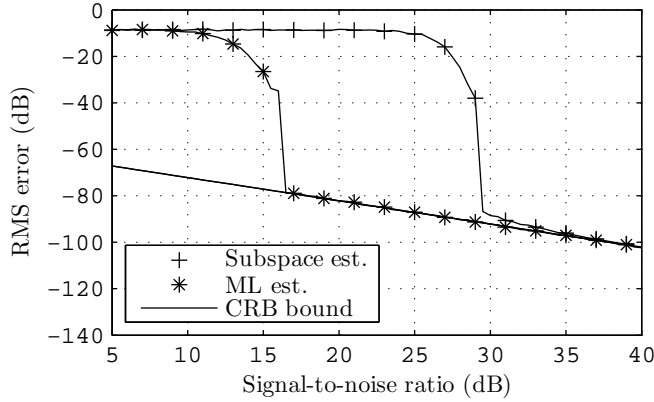\end{aligned}
\tag{26}
$$

Fig. 3. Root-Mean-Square (RMS) error of the subspace and ML estimators, together with the Cramer-Rao bound.
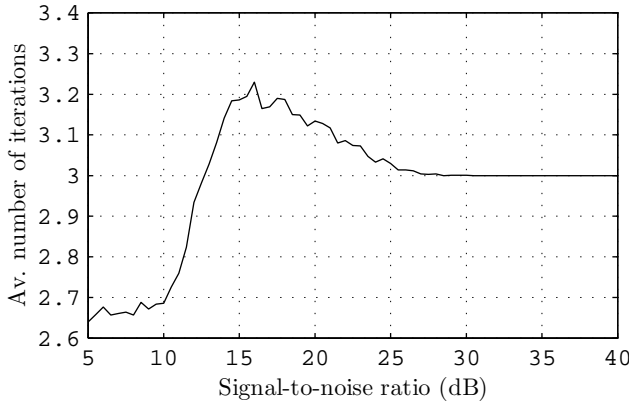

Fig. 4. Average number of iterations required by the Newton method.

The actual estimations $\hat{f}_{o,1}$ and $\hat{f}_{o,2}$ were computed from $\mathbf{u}_1$ and $\mathbf{v}_1$ by applying the 1-D method in Sec. 3 to each of them. This method had complexity $O(MN(M + N))$, since it involved the computation of the singular vectors $\mathbf{u}_1$ and $\mathbf{v}_1$, and is equivalent to the method in [3]. The second method computed the actual ML estimator from $\mathbf{Z}$ using the method in Sec. 4, and its complexity was just $O(MN \log(MN))$. For the values of M and N given above, the over-sampling factors were 2.048 and 1.523, respectively, i.e, the FFT had size 1024 in both dimensions.

Fig. 3 shows the root-mean-square (RMS) error for the first estimator, termed subspace estimator, for the second estimator (interpolated ML estimator), and the Cramer-Rao bound. The threshold performance of the second is clearly superior.

Finally, Fig. 4 shows the average number of iterations that required the interpolated ML estimator, which was roughly equal to three.

6. CONCLUSIONS

A method has been presented that allows one to compute the maximum likelihood (ML) estimation of a complex 2-D sinusoidal, with the complexity of the fast Fourier transform (FFT). First, it is recalled in the paper that a band-limited signal can be interpolated with high accuracy from a small number of samples, if the sampling frequency is somewhat higher than the Nyquist frequency. Besides, a specific barycentric formula is proposed to perform this kind of interpolation. And second, it is shown that the ML cost function for the estimation of a complex 2-D (and 1-D) sinusoidal can be viewed as a band-limited signal, if the time and frequency variables are switched. Finally, these two results are combined in a method that is able to deliver the ML estimate with the complexity order of the FFT, which is based on Newton's algorithm.

7. REFERENCES

[1] B. G. Quinn, "A fast efficient technique for the estimation of frequency," *Biometrika*, vol. 78, no. 3, pp. 489–497, 1991.

[2] B. G. Quinn, "Estimating frequency by interpolation using Fourier coefficients," *IEEE Transactions on Signal Processing*, vol. 42, no. 5, pp. 1264–1268, 1994.

[3] H. C. So, F. K. W. Chan, W. H. Lau, and C. Chan, "An efficient approach for two-dimensional parameter estimation of a single-tone," *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 1999–2009, Apr 2010.

[4] E. Aboutanios and B. Mulgrew, "Iterative frequency estimation by interpolation on Fourier coefficients," *IEEE Transactions on Signal Processing*, vol. 53, no. 4, pp. 1237–1242, Apr 2005.

[5] M. D. Macleod, "Fast nearly ML estimation of the parameters of real or complex single tones or resolved multiple tones," *IEEE Transactions on Signal Processing*, vol. 46, no. 1, pp. 141–148, Jan 1998.

[6] I. Perisa and J. Lindner, "Employing simple FFT-interpolation for improved complex tone detection and fine estimation," in *3rd International Symposium on Wireless Communication Systems, 2006. (ISWCS '06)*, 6-8 2006, pp. 744 –748.

[7] T. I. Laakso, V. Välimäki, M. Karjalainen, and U. K. Laine, "Splitting the Unit Delay," *IEEE Signal Processing Magazine*, vol. 13, no. 1, pp. 30–60, Jan 1996.

[8] J. Selva, "Design of barycentric interpolators for uniform and nonuniform sampling grids," *IEEE Transactions on Signal Processing,*, vol. 58, no. 3, pp. 1618 –1627, Mar 2010.